

09/242772

PCIP

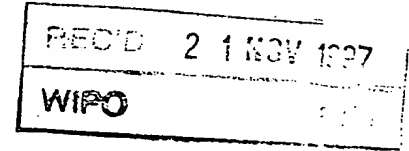
97/04759



Eur päisches  
Patentamt

European  
Patent Office

Office européen  
des brevets



Bescheinigung

Certificate

Attestation

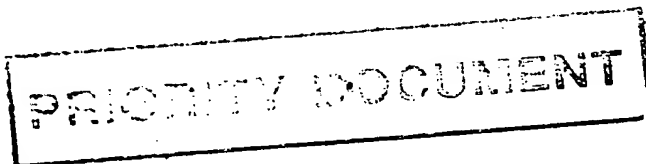
Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

96202339.6



Der Präsident des Europäischen Patentamts:  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

Ch. D. ...

DEN HAAG, DEN  
THE HAGUE,  
LA HAYE, LE

11/97



Eur päisches  
Patentamt

Eur pean  
Patent Office

Office eur péen  
des brevets

**Blatt 2 der Bescheinigung**  
**Sheet 2 of the certificate**  
**Page 2 de l'attestation**

Anmeldung Nr  
Application no  
Demande n° 96202339.6

Anmeldetag:  
Date of filing: 22/08/96  
Date de dépôt:

Anmelder  
Applicant(s)  
Demandeur(s)  
K. U. LEUVEN RESEARCH & DEVELOPMENT  
B-3000 Leuven  
BELGIUM

Bezeichnung der Erfindung  
Title of the invention  
Titre de l'invention

Promoter swapping between the genes for a novel zinc finger protein and B-catenin in pleomorphic adenoma of the salivary glands

In Anspruch genommene Priorität(en) / Priorities claimed / Priorité(s) revendiquée(s)

Staat  
State  
Pays

Tag  
Date  
Date

Aktenzeichen  
File no  
Numero de dépôt

Internationale Patentklassifikation  
International Patent classification  
Classification internationale des brevets

C12N15/00

Am Anmeldetag benannte Vertragsstaaten  
Contracting states designated at date of filing: AT/BE/CH/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE  
Etats contractants désignés lors du dépôt:

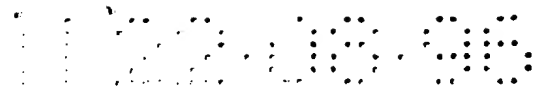
Bemerkungen  
Remarks  
Remarques

14-22-00-95

2

**PROMOTER SWAPPING BETWEEN THE GENES FOR A NOVEL ZINC  
FINGER PROTEIN AND  $\beta$ -CATENIN IN PLEOMORPHIC ADENOMA OF THE SALIVARY GLANDS**

## INTRODUCTION



Recently, molecular insight was obtained into the genetic basis of benign tumors by the discovery of a common genetic denominator in tumors of different types but with a particular cytogenetic profile<sup>1,2</sup>. Molecular analysis of recurrent chromosome 12q13-15 aberrations in pleomorphic adenoma of the salivary glands, lipoma, uterine leiomyoma, hamartoma of lung and breast, fibroadenoma of the breast, angiomyxoma, and endometrial polyps revealed the consistent involvement of the high mobility group protein gene *HMGIC*; the first "benign oncogene". Evidence is mounting that the same is likely to hold true for the closely related family member *HMGII*, which is located on the short arm of chromosome 6. Both genes belong to the high mobility group (HMG) protein gene family<sup>3</sup>. Their corresponding proteins possess three amino-terminal AT hooks, through which the proteins are assumed to bind to A/T-rich DNA sequences, and an acidic tail in the carboxy-terminal region. Functionally, they act as architectural factors in the nuclear scaffold, critical for the correct assembly of stereospecific transcriptional complexes<sup>4</sup>. The genetic aberrations frequently result in the separation of the three AT hooks from the acidic tail, although all the pivotal changes in the high mobility group protein genes that can lead to aberrant growth control are not yet known. It should be noted that a variety of chromosome segments can act as translocation partner of 12q13-15 but each tumor type seems to have a preferential translocation partner. At present, only the preferential t(3;12)(q27-28;q15) translocation of ordinary lipoma has been characterized<sup>5</sup>. The gene affected on chromosome 3 is the *LPP* gene, which encodes a protein containing three LIM domains. LIM domains have been found in a growing number of proteins in mammals, amphibians, flies, worms, and plants and act as modular protein-binding interfaces<sup>6</sup>. LIM proteins mainly have a function in cell signalling and developmental regulation and include, for instance, transcription regulators, proto-oncogene products, and adhesion plaque constituents. The preferential t(3;12) in lipoma results in the formation of an *HMGIC/LPP* fusion transcript encoding a hybrid protein consisting of the three DNA binding domains of *HMGIC* and the LIM domains of the protein encoded by *LPP*.

In an approach to systematically unravel the molecular pathogenesis of benign tumors, we have focused on other cytogenetic subgroups of these, since their mere existence indicate the presence of additional critical genes. It is of interest to establish whether such genes are functionally similar or different as compared to the high mobility group protein genes *HMGIC* and *HMGII*. In this context, we have now molecularly characterized the largest cytogenetic subgroup of pleomorphic adenomas. This type of neoplasm is an epithelial tumor occurring primarily in the major and minor salivary glands. It is by far the most common type of salivary gland tumor, accounting for almost 50 % of all tumors in these organs<sup>7</sup>. Pleomorphic adenomas are almost exclusively benign tumors, which only rarely

undergo malignant transformation. They show a marked histological diversity with epithelial and myoepithelial cells arranged in a variety of patterns in a matrix of mucoid, myxoid, chondroid and, on rare occasions, even osteoid tissues. Cytogenetically, pleomorphic adenomas are characterized by recurrent rearrangements, in particular reciprocal translocations with consistent breakpoints at 3p21, 8q12 and 12q13-15 (ref. 8). In addition to the cases with abnormal karyotypes, there is also a subgroup of tumors with apparently normal karyotypes (30 to 50 % of the cases). Abnormalities of 8q12 are most common and are found in about 60 % of the cases with abnormal stemlines, whereas rearrangements of 3p21 and 12q13-15 are found in about 30 % and 20 % of the cases, respectively. The most frequent and characteristic abnormality so far observed in pleomorphic adenomas is a reciprocal  $t(3;8)(p21;q12)$  translocation<sup>9</sup>. While chromosome 3p is the preferential translocation partner, most human chromosomes have been found as translocation partner of 8q12. Rearrangements of 8q12 are often found as the sole anomalies, indicating that they represent primary cytogenetic events of possible pathogenetic importance.

Here, we describe the positional cloning of the 8q12 translocation breakpoint as well as the identification and characterization of the genes on chromosome 8q12 and 3p21 disrupted by the  $t(3;8)$  translocation. The gene on chromosome 8q12 is a novel zinc finger gene, which we have designated *PLAG1*, whereas the gene on 3p21 is *CTNNB1*, the gene for  $\beta$ -catenin, a protein with an established role in cell adhesion and signal transduction<sup>10</sup>. By Southern blot and RACE analysis, we pinpoint the breakpoints within the introns of the 5'-noncoding regions of both genes and show that well-defined fusion transcripts are expressed in the tumor cells. We also demonstrate that the  $t(3;8)$  translocation results in specific disruption of the transcriptional control of the two genes, leading to activation of *PLAG1* and down-regulation of *CTNNB1*. A pleomorphic adenoma with 8q12 but not 3p21 involvement has also been studied to evaluate the phenomenon of disruption of transcriptional control in a broader perspective.

## RESULTS

### The chromosomal 8q12 breakpoint cluster region in pleomorphic adenoma

As part of a positional cloning effort to isolate the gene(s) affected by the t(3;8)(p21;q12) in pleomorphic adenomas, we constructed a yeast artificial chromosome (YAC) contig consisting of 34 overlapping YACs, spanning about 2 Mb within band 8q12 (to be published elsewhere). A long range STS and rare cutter physical map of this region was also developed. In previous experiments the t(3;8) breakpoint was mapped to a 1 Mb region flanked by *MOS* as proximal and by D8S166 as distal marker. One YAC within this region was shown to span the t(3;8) breakpoint in two tumors (CG588 and CG644)<sup>11</sup>. To further narrow down the breakpoint region we isolated cosmids corresponding to landmarks within this and other YACs in our contig (Fig. 1A). By FISH, we could demonstrate that the breakpoints in four out of four adenomas tested (CG580, CG588, CG644 and CG682) clustered within a 300 kb subregion located between *MOS* and STS CH129 (Fig. 1A). Two cosmids from this region, CEM23 and CEM48, were shown to span the breakpoints in adenomas CG644 (Fig. 1B) and CG682 (data not shown), respectively. To refine the distribution of breakpoints within this 300 kb subregion, we developed a contig consisting of 29 phage and cosmid clones (Fig. 1A).

### Identification and characterisation of the *PLAG1* gene in the 8q12 breakpoint region

While initiating experiments to identify candidate genes mapping within the 300 kb genomic segment harbouring the adenoma breakpoints, BLAST searches<sup>12</sup> of newly obtained STSs mapping within the 300 kb contig revealed that the right end (CH283) of YAC clone 143D5 displayed sequence identity with a publicly available EST (expressed sequence tag) (GenBank Accession number D59273). Its position in the contig raised the possibility that this EST belonged to the candidate pleomorphic adenoma gene (*PLAG1*), we were searching for. In initial Northern blot experiments using a PCR probe corresponding to this particular YAC end, a 7.5 kb transcript was detected. A combination of sequential screenings of a human fetal kidney cDNA library as well as 5'- and 3'-RACE experiments led to the isolation of a composite cDNA of 7313 nucleotides (GenBank accession number XXX). This cDNA contains an open reading frame (ORF) of 1500 bp starting with the ATG at position 481-483 (Fig. 2A). An in frame stop codon (TAG) is present 9 nucleotides upstream of this ATG. The deduced amino acid sequence reveals seven canonical C2H2 zinc finger domains (Fig. 2B) and a non-finger region of 259 amino acid residues representing the carboxy-terminus of the deduced protein. The zinc finger motifs including the linker sequences are between 28 and 35 amino acids long. The cysteine (C) and histidine (H) residues are present in their characteristic positions in each finger. The

typical phenylalanine (F) and leucine (L) residues in finger 4 and 6 are lacking. Strictly spoken, the deduced protein is not a Kruppel zinc finger protein, since it does not contain the characteristic H/C linker (consensus sequence TGEKPYK) in between the zinc fingers<sup>13</sup>. Only the seven amino acids between finger 1 and 2 resemble the H/C linker (TGERPYK). The amino-terminal region contains two nuclear localization signals (KRKR and KPRK). The carboxy-terminus is serine-rich (45 amino acid residues out of 259, i.e 17%), raising the possibility of a regulatory role that may be controlled by serine/threonine kinases.

Outside the ORF, some interesting features in the cDNA can be observed. In the 3' untranslated region (UTR), which is 5333 bp long, a polyadenylation signal is present starting at position 7297, and there is also a TG-repeat, which might be of regulatory relevance. In addition, there is an ATTTA sequence motif, which has previously been shown to mediate mRNA destabilization in certain lymphokine and immediate early genes<sup>14</sup>. The ATTTA motif is repeated twelve times in the 3'-UTR.

No sequence similarity was found to any of the known zinc finger proteins but preliminary studies have indicated that the human genome contains at least one additional gene that is closely related to our candidate pleomorphic adenoma gene. Nine anonymous ESTs (expressed sequence tags) with sequence similarity to sequences of our new gene were found in public databases and these ESTs could be aligned resulting in a contiguous sequence of about 1 kb. An open reading frame appeared to be present with coding potential for protein sequences with zinc-finger-like features. Sequence identity between the aligned ESTs and our new gene appeared to be about 75% at the amino acid sequence level; i.e. in the region of our gene encoding zinc fingers 4 to 7. Of interest to note furthermore is that these anonymous ESTs sequences have been mapped to chromosome 6q24, which is implicated in tumors of the salivary glands.

#### **Genomic organization of the *PLAG1* gene**

Comparison of transcribed and genomic DNA sequences of the *PLAG1* gene revealed that it contains 5 exons (Fig. 1A). The transcriptional orientation of the gene is directed towards the centromere. The first three exons are noncoding, the fourth exon contains the translation start site (ATG), and the amino-terminus of the protein including one complete zinc finger domain. The second finger is split by intron 4 and continues into exon 5, which contains the remaining part of the ORF and the long 3'-UTR (5533 bp). Based on our YAC and cosmid maps, the *PLAG1* locus spans about 35 kb, with a large intron (approximately 25 kb) between exon 1 and exon 2. 3'-and 5'-RACE analysis revealed the existence of two alternatively spliced mRNAs which differ from each other by the presence or absence of the noncoding exon 2. The difference in electrophoretic mobility of the two mRNA isoforms is too small to

distinguish them by Northern blot analysis. The functional relevance of this alternative splicing remains to be elucidated.

### **Rearrangements of *PLAG1* in primary pleomorphic adenoma**

In an attempt to establish whether the 8q12 aberrations in primary pleomorphic adenomas with t(3;8)(p21;q12) led to rearrangements in *PLAG1*, Southern blot experiments were performed using various *PLAG1*-derived probes. With a probe within the ORF (STS probe EM265) or the 3'-UTR (KK64) of *PLAG1*, no rearrangements were detected. However, using probes corresponding to the 5' noncoding region of *PLAG1*, i.e. intron 1 (STS probes EM416 or EM317) or exon 2 (STS probe EM440), rearrangements were detected in each of the t(3;8) tumors tested (data not shown). Similar Southern blot analysis of primary pleomorphic adenoma CG580, which carries a t(8;15)(q12;q14), revealed also a rearrangement in the 5' noncoding region of *PLAG1* (Fig. 3). These results seem to directly implicate *PLAG1* as a critical gene in pleomorphic adenoma tumorigenesis. Furthermore, the mechanism might involve transcriptional deregulation of *PLAG1*, since the rearrangements in *PLAG1* were invariably found in the 5' noncoding region of *PLAG1* leaving its coding region intact.

### **t(3;8) results in promoter swapping between *PLAG1* and the gene for $\beta$ -catenin, *CTNNB1***

The observation that the 8q12 translocation breakpoints in various primary tumors were mapped to intron 1 or intron 2 of the 5'-noncoding region of *PLAG1* raised the possibility that the t(3;8) results in the production of a chimeric transcript consisting of *PLAG1* sequences fused to those of a gene located on chromosome 3p21. To test this possibility, 5'-RACE experiments were performed using total RNA of the primary tumors CG644 and CG682, which both carry a 3;8-rearrangement. We used a primer specific for exon 5 (MV2) in the cDNA synthesis. Following ligation of an adaptor to the cDNA, PCR amplification was performed using, in the first round, an adaptor-specific primer and a *PLAG1*-specific primer corresponding to sequences of exon 5 (MV5) and, in a second round, nested adaptor-specific primer and *PLAG1*-specific primer corresponding to sequences of exon 4 (MV6). Nucleotide sequence analysis of the PCR product revealed that the ectopic sequences were fused to the acceptor splice site of exon 3 of *PLAG1*. BLAST analysis revealed that they were identical to exon 1 sequences of *CTNNB1*<sup>15</sup>, the gene for  $\beta$ -catenin, which has previously been assigned to chromosome 3p21 (ref. 16).

Independent evidence for the involvement of *CTNNB1* in the t(3;8)-rearrangements in pleomorphic adenomas was obtained by FISH using two YACs (756G5 and 750D3), which contain the complete gene for  $\beta$ -catenin. According to published maps YAC 756G5 is contained within the proximal part of mega-YAC 750D3 (ref. 17). As expected, both

YACs spanned the 3p21 breakpoint in adenoma CG682, with hybridization signals on the normal 3, the der(3) and the der(8) (data not shown). Detailed analysis of a few prometaphase cells enabled us to sublocalize the  $\beta$ -catenin locus to band 3p21.3, which is in accordance with previous mapping data.

To extend our observation to a larger number of tumors, we applied a reverse transcription-polymerase chain reaction (RT-PCR) approach. RT-PCR amplifications using primers specific for *CTNNB1* and *PLAG1* were carried out with RNA from tumors CG368, CG588, CG644, CG752, CG753 and T9587, which all carry the recurrent t(3;8), and from tumor CG682, which carries an ins(8;3)(q12;p21.3p14.1). RNA from tumor CG580, which carries a t(8;15), was included as a negative control. PCR experiments resulted in the generation of PCR products corresponding to hybrid transcripts consisting of *PLAG1* and *CTNNB1* sequences, in seven out of seven t(3;8) tumors analyzed (Fig. 4). In tumors CG368, CG588, CG682, CG752, and T9587, PCR products of 509 bp (Fig. 4A, PCR product A) and 614 bp (Fig. 4A, PCR product B) were generated, whereas in tumors CG644 and CG753, only the PCR product of 509 bp was found. The PCR product of 509 bp (from NECAT-UP up to MV6) corresponds to a hybrid transcript containing exon 1 of *CTNNB1* and exons 3 to 5 of *PLAG1*. The PCR product of 605 bp contains an extra 105 bp, which corresponds to the alternatively spliced exon 2 of *PLAG1*. It points towards the presence of a related isoform consisting of exon 1 of *CTNNB1* and exons 2 to 5 of *PLAG1*. This was also confirmed by nucleotide sequence analysis of the PCR products. We were also able to demonstrate that the corresponding reciprocal fusion transcripts are expressed. In all tumors except CG682, a PCR product of 130 bp was generated corresponding to a fusion transcript consisting of exon 1 of *PLAG1* and exons 2 to 16 of *CTNNB1*. In tumor CG753, an additional PCR product was detected, corresponding to a fusion transcript consisting of exons 1 to 2 of *PLAG1* and exons 2 to 16 of *CTNNB1*. This additional band was also observed but with weak intensity in tumor CG644. All these results indicate that in CG644 and CG753, the 8q12 translocation breakpoints are located in Intron 2, whereas the breakpoints of tumors CG368, CG588, CG682, CG752, and T9587 are located in Intron 1. Interestingly, using 5' RACE analysis with the tumor CG580, which carries a t(8;15) translocation, we found that the breakpoint occurs also in the same region leading to a fusion transcript with ectopic fused to exon 3 of *PLAG1*.

#### **Activation of *PLAG1* expression due to promoter swapping**

Northern blot analysis was performed to evaluate the effect of the t(3;8)(p21;q12) translocation on the expression levels of *PLAG1* and *CTNNB1* (Fig. 5). As previously mentioned, *PLAG1* is expressed as a 7.5 kb transcript, which was readily detected in human fetal lung, liver, and kidney but not in fetal brain. In adult tissues, the 7.5 kb transcript was not detected in human heart, brain, lung, liver, skeletal muscle, kidney, pancreas, and salivary gland.

Low levels of *PLAG1* expression was found in human placenta (Fig. 5A). In contrast, the *CTNNB1* gene was ubiquitously expressed as a 3.8 kb RNA doublet in all tissues tested (Fig. 5B). Similar results were obtained in the Northern evaluation of fetal and adult mouse tissues (data not shown).

Northern blot analysis was performed on two pleomorphic adenomas with t(3;8) and on one tumor with a variant t(8;15) (Fig. 5C). In contrast to normal salivary gland tissue in which no *PLAG1* transcripts could be detected, the three tumors expressed a 7.5 kb transcript. This transcript does not correspond to the normal *PLAG1* transcript (exons 1 to 5), since an exon 1-specific probe failed to hybridize to this transcript (data not shown). Instead, a probe specific for exon 1 of *CTNNB1* recognized this 7.5 kb transcript as well as the 3.8 kb doublet of *CTNNB1* transcripts in the two tumors with a t(3;8). In contrast, only the 3.8 kb doublet of *CTNNB1* transcripts was detected in adenoma CG580 with a variant t(8;15). The expression level of *CTNNB1* in this tumor was about twofold higher as compared to tumors CG644 and CG682, which is expected since none of the *CTNNB1* alleles is rearranged. Collectively, these results suggest that *PLAG1* expression in these pleomorphic adenomas is driven by the 5' regulatory sequences either of the *CTNNB1* gene for the cases with t(3;8) or of an as yet unknown gene on chromosome 15 for the case with t(8;15). These results indicate that the principal mechanism in pleomorphic adenomas with 8q12 rearrangements is activation of *PLAG1* expression, due to promoter swapping between *PLAG1* and a translocation partner gene, preferentially *CTNNB1*. Furthermore, expression of the translocation partner gene is likely to be down-regulated concomitantly; e.g. as demonstrated for *CTNNB1*.

## DISCUSSION

11-20-05-95

We have characterized a novel gene, *PLAG1*, which we propose is a critical locus involved in the development of pleomorphic adenoma of the salivary glands. The gene was identified on the long arm of chromosome 8, close to the *MOS* proto-oncogene, as a result of a positional cloning project to molecularly define the t(3;8)(p21;q12), which is the most frequent chromosome aberration in pleomorphic adenomas. We have also identified the translocation partner gene at 3p21, i.e. the *CTNNB1* gene which encodes  $\beta$ -catenin. In addition to 8q12 and 3p21, chromosome rearrangements involving 12q13-15 are also frequently observed in pleomorphic adenomas. The high mobility group protein gene *HMGIC* was recently identified as the gene consistently rearranged in these cases<sup>1</sup>. In conclusion, we have molecularly defined the three main chromosomal rearrangements implicated in the genesis of pleomorphic adenomas.

Structural rearrangement of the 8q11-13 region are consistently found also in benign lipoblastomas. These tumors are considered to be unique variants of lipoma, occurring primarily in children before 3 years of age. Among the seven cases so far analyzed cytogenetically, rearrangements involving 8q11-13 were observed in several cases<sup>8</sup>. Similar to pleomorphic adenomas, the translocation partners of 8q12 are variable. The similarities in chromosomal pattern between these two tumor types raise the question as to whether *PLAG1* is the target of the 8q11-13 rearrangements also in lipoblastomas. Molecular cloning of the translocation breakpoints, which are now in progress, will resolve this question.

We have also elucidated the molecular consequences of the t(3;8) translocation. The 8q12 translocation breakpoints were invariably found in the 5'-non-coding region of *PLAG1* in all seven primary tumors tested. The breakpoints in the *CTNNB1* gene were always found in the first intron. As a result of the translocation, the coding sequences of *PLAG1* are brought under control of the 5' regulatory sequences of the *CTNNB1* gene, and vice versa, resulting in the activation of *PLAG1* and down-regulation of *CTNNB1*. Based on their expression patterns in normal fetal and adult tissues, it is clear that the regulation of expression of the two genes is very different. The *CTNNB1* gene is highly and ubiquitously expressed, whereas *PLAG1* is developmentally regulated, with expression restricted to fetal tissues. It should be emphasized that in adult tissues, *PLAG1* is either not expressed or expressed at very low levels. In the adenomas, both the *CTNNB1/PLAG1* and the reciprocal *PLAG1/CTNNB1* transcripts were detected by RT-PCR; detection of the latter transcripts sometimes required three rounds of PCR, indicating that the expression levels are very low. Our findings represent the first example of reciprocal exchange of expression control elements in solid tumors. Since the coding sequences of both genes are invariably preserved, the molecular mechanism could be

classified as promoter swapping, resulting in activation of *PLAG1*, and down-regulation of *CTNNB1* expression. Promoter substitutions have previously been described in B-cell lymphomas and T-cell leukemias, the best known examples comprising rearrangements involving immunoglobulin (Ig) and T-cell receptor (TCR) genes<sup>18,19</sup> and references therein. Examples without Ig or TCR gene involvement are scarce. The B-cell lymphoma cell line VAL bears a t(3;4)(q23;p11) translocation resulting in the expression of a chimeric RNA transcript between the *LAZ3/BCL6* gene and the 5' UTR of *TTF*, a small G protein encoding gene<sup>20</sup>. In the B-cell leukemia cell line Karpas 231 the entire coding part of *LAZ3/BCL6* is placed under the control of the *BOB1/OBF1* 5' regulatory sequences<sup>21</sup>. However, in none of these cell lines, a reciprocal chimeric fusion transcript corresponding to the 5' regulatory region and first exon of *LAZ3/BCL6* fused to the *TTF* or *BOB1/OBF1* sequences, could be detected, neither by Northern blot nor by RT-PCR. A similar promoter substitution has also been described in T-cell acute lymphoblastic leukemia (ALL) where a 90 kb deletion results in a head-to-tail fusion of two genes, *SIL* and *SCL*, leading to *SCL* mRNA expression regulated by the *SIL* promoter and thus inappropriate *SCL* expression<sup>22</sup>.

Since the t(3;8) in pleomorphic adenomas invariably leads to activation of *PLAG1*, it is of interest to evaluate possible physiological implications. The *PLAG1* gene encodes a protein with a deduced molecular weight of 56 kDa and a deduced pI of 8.56. Analysis of the open reading frame of *PLAG1* reveals seven zinc fingers in the amino-terminal region. The carboxy-terminal region is rich in serine residues. Furthermore, two potential nuclear localization signals are present (residues 22-25 and 29-32). Collectively, this suggests that the *PLAG1* protein is a novel member of the large zinc finger gene family. Zinc finger motifs were originally identified as DNA binding structures in the RNA polymerase III transcription factor TFIIIA, which binds to the internal control region of the 5S RNA gene. TFIIIA-like zinc fingers are present in a variety of regulatory proteins found in higher and lower eukaryotes. This type of zinc finger motif consists of ~30 amino acids with two cysteine and two histidine residues (C2H2) that stabilise the domain by tetrahedrally coordinating a  $Zn^{2+}$  ion. A region of ~12 amino acids between the invariant cysteine-histidine pairs is characterized by scattered basic residues and several conserved hydrophobic residues. Most zinc finger genes studied so far encode polymerase II transcription factors. Apart from transcriptional modulation and control of RNA metabolism, chromatin packaging might also constitute an important activity through which zinc finger proteins exert their regulatory roles<sup>23</sup>. It should also be noted that the presumed role of *HMGIC* is also in chromatin modelling<sup>1,2,4</sup>. The mammalian genome is known to contain a large number of C2H2 zinc finger genes<sup>13</sup> and the number of such genes implicated in cancer is growing steadily. For example, mutations within the zinc finger gene *WT1* at 11p13 predispose to Wilms' tumor<sup>24</sup>. Furthermore, in desmoplastic small round cell tumors, *WT1* is fused to the *EWS* gene on chromosome 22q12 as the result of a t(11;22) (ref. 25), and in acute promyelocytic leukemia, a fusion occurs

between the retinoic acid receptor locus and a Kruppel-like zinc finger gene *PLZF*, due to a t(15;17), or to the zinc finger gene *PLZF*, due to a variant t(11;17) translocation<sup>29,27</sup>. Similarly, in non-Hodgkin's lymphoma, *LAZ3/BCL6*, a gene encoding a protein with seven zinc fingers, is rearranged by recurrent translocations involving 3q27<sup>28,29</sup>. The function of the serine rich carboxy-terminal part of *PLAG1* is unknown but it may have a regulatory function that can be controlled by serine/threonine kinases. If *PLAG1* encodes a DNA binding protein, as the presence of its zinc fingers suggests, the carboxy-terminal region might represent the transactivation domain. At least four different primary sequence motifs that characterize the activation domains are identified thus far, i.e. acidic, glutamine-rich, proline-rich and serine/threonine-rich. They likely represent regions that functionally interact with other proteins<sup>30</sup>.

Activation of *PLAG1* due to promoter substitution may lead to deregulation of genes normally controlled by *PLAG1*. If overexpression of *PLAG1* is indeed a major molecular consequence of the 8q12 translocations than this could possibly be achieved also by other mechanisms, including for instance an increase in gene copy number. In this context it should be noted that a small subgroup of pleomorphic adenomas show trisomy 8, often as the sole anomaly<sup>8</sup>. In addition, trisomy 8 is a common numerical abnormality found in several histological subtypes of malignant salivary gland tumors as well as in different subtypes of leukemias and sarcomas<sup>8</sup>. Whether trisomy 8 may lead to overexpression of *PLAG1* remains, however, to be shown.

The preferential fusion partner of *PLAG1* is the *CTNNB1* gene, which encodes  $\beta$ -catenin, a cytoplasmic protein of about 88 kDa<sup>31</sup>. Its frequent involvement in pleomorphic adenomas as found here might point towards a critical role of *CTNNB1*. Most likely, a role of *CTNNB1* is to provide an active promoter in front of the *PLAG1* gene. However, since the observed promoter swapping between *PLAG* and *CTNNB1* leads to down-regulation of *CTNNB1* expression, it may also lead to other physiological consequences, especially since  $\beta$ -catenin is a broad-range protein interface, as that has been implicated in highly diverse processes, such as human colon cancer, epithelial cell adhesion, embryonal axis formation in *Xenopus*, and pattern formation in *Drosophila*<sup>10</sup>. The  $\beta$ -catenin protein has also been found as structural component of adherens junctions (AJs)<sup>10,32,33</sup>, which are multiprotein complexes assembled around  $\text{Ca}^{2+}$ -regulated cell adhesion molecules (cadherins).  $\beta$ -Catenin binds directly to cadherins and acts as a protein interface between cadherin and the cytoskeleton. The cadherin- $\beta$ -catenin complex mediates cell adhesion, cytoskeletal anchoring, and signalling, which are important processes for regulation of cell growth and behaviour. It has been suggested that AJ complexes may play a role in the transmission of signals for contact inhibition<sup>10</sup>. Down-regulation of  $\beta$ -catenin, as occurs in pleomorphic adenomas, could interfere with this transmission and lead to growth deregulation. In other studies,  $\beta$ -catenin has been linked to processes implicated in polyp formation, an early stage in colorectal tumorigenesis, since it has been found to associate with the adenomatous polyposis coli protein APC<sup>34,35</sup>. APC is a

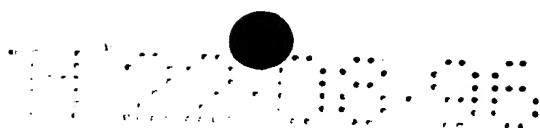
large protein which in addition to  $\beta$ -catenin binding regions, contains multiple domains including an oligomerization domain, multiple protein-protein interaction motifs (Arm repeats), a basic domain and a domain that can bind to microtubules<sup>36</sup>. Inheritance of one mutant APC allele is known to result in predisposition to colorectal cancer. Very recently, evidence was provided for a firm place of APC in the WG/WNT signalling pathway<sup>37,38</sup>; WG (WINGLESS) is a cell-to-cell signal in *Drosophila* that triggers several key developmental processes and WNT is the analogous mouse molecule. The role of APC in the WG/WNT signalling pathway is not yet fully clear but two models have been suggested that include complex formation with  $\beta$ -catenin<sup>38</sup>; APC may act solely as a negative regulator of  $\beta$ -catenin accumulation and signalling, or APC may act both as a negative regulator and effector, with phosphorylation of APC by protein kinase GSK3 $\beta$  as controlling element. Phosphorylated APC may down-regulate intracellular  $\beta$ -catenin keeping signalling off, whereas WG/WNT can antagonize protein kinase GSK3 $\beta$ , stabilizing intracellular  $\beta$ -catenin and activating signalling. Since  $\beta$ -catenin levels apparently are important physiologically, it is tempting to speculate that down-regulation of *CTNNB1* expression might play some role in pleomorphic adenoma development, possibly synergistically to *PLAG1*. In conclusion, as activation of *PLAG1* is observed in all pleomorphic adenomas with 8q12 involvement tested, this seems to constitute a major factor in salivary gland tumor formation. The critical role of down-regulated levels of  $\beta$ -catenin remains to be established.

Previous studies of *HMGIC* have shown that rearrangements of this gene preferentially occur in particular cytogenetic subgroups of benign mesenchymal tumors, with the only exception so far being a subgroup of pleomorphic adenomas<sup>1</sup>. When considering the findings in the latter type of tumors, it should be noted that, although they are epithelial in origin, they often also show a varying degree of mesenchymal differentiation. This may be explained by the fact that pleomorphic adenomas are thought to originate from a pluripotent reserve cell in the terminal duct system which possesses the capacity to differentiate into both epithelial and myoepithelial cells<sup>39</sup>. The latter cells may function as facultative mesenchymal cells, and are responsible for the production of extracellular material, including myxo-chondroid stroma<sup>40</sup>. An intriguing question in this context is of course whether *HMGIC* is preferentially affected in adenomas with a prominent stromal component while *PLAG1* is preferentially affected in tumors with little or no stroma. Histological re-examination of all but one of the adenomas with *PLAG1* involvement in this study, revealed that all cases contained mesenchymal components, including both myxoid and chondroid tissues, suggesting that mesenchymal differentiation is not restricted to adenomas with *HMGIC* involvement. It should be noted that previous cytogenetic observations have already indicated that chromosome 12q13-15 and 8q12 rearrangements constitute mutually exclusive pathways for the evolution of pleomorphic adenomas<sup>41</sup>.

The identification and cloning of a novel "benign oncogene" activated by chromosome translocations constitute

a major step towards an increased understanding of the molecular pathogenesis of benign neoplastic growth. Our findings may lead to new insights into the genetic differences between benign and malignant tumors. In addition, the present findings may have diagnostic implications. The fact that *PLAG1* is not normally expressed in adult tissues, but activated in pleomorphic adenomas due to 8q12 rearrangements, makes it a potentially useful tumor marker in the differential diagnosis of benign and malignant salivary gland tumors. On the basis of the fusion transcripts that are formed, highly specific diagnostic assays can be developed.

## METHODS



### Tumor material and chromosome analysis

Primary pleomorphic adenomas of the salivary glands were obtained from patients at the time of surgery. Primary cultures and chromosome preparations were made and analyzed as described by R ijer *et al.*<sup>11</sup>. Primary pleomorphic adenomas, including CG368, CG580, CG644, CG752, CG753 and T9587, which all carry the recurrent t(3;8)(p21;q12) as the sole anomaly, were selected for molecular analysis, as well as CG682, showing an ins(8;3)(q12;p21.3p14.1)<sup>11</sup>, and CG588 which carries a t(8;15)(q12;q14). FISH analysis was performed as previously described<sup>11</sup>. Slides were examined in a Zeiss Axiophot epifluorescence microscope using the appropriate filter combinations. Fluorescence signals were digitalized, enhanced and analyzed using the ProbeMaster FISH image analysis system (Perceptive Scientific Instruments, Houston, Texas). Color prints were produced using a Kodak XL 7700 monochrome continuous printer.

### Preparation and analysis of DNA and RNA

Extraction of genomic DNA and Southern blot analysis were performed as described previously<sup>42</sup>. Total RNA was extracted from biopsy samples using the TRIZOL (Gibco/BRL) method and Northern analysis was performed according to standard procedures<sup>43</sup>. DNA from YACs, cosmids, PCR products, and oligonucleotides was labelled using a variety of techniques. For FISH, cosmid clones or inter-Alu PCR products of YACs were biotinylated with biotin-11-dUTP (Boehringer) by nick translation. For filter hybridizations, probes were radio-labelled with  $\alpha$ -<sup>32</sup>P-dCTP using random hexamers<sup>44</sup>. In case of PCR-products smaller than 200 bp in size, a similar protocol was applied, but specific oligonucleotides were used to prime labelling reactions. Oligonucleotides were labelled using  $\gamma$ -<sup>32</sup>P-ATP.

### YAC, cosmid, phage, and cDNA libraries

YAC clones in this paper were isolated from the CEPH mark 1 YAC library<sup>45</sup>, using a combination of PCR-based screening and colony hybridization analysis<sup>46</sup>. YAC DNA was isolated and characterized as described before<sup>1,42</sup>. Cosmid clones were isolated from an arrayed human chromosome 8-specific cosmid library<sup>47</sup> obtained from Los Alamos National Laboratory (LANL). LANL-derived cosmid clones are indicated by their unique microtiter plate addresses. Phage clones were derived from a genomic library constructed with Li-14/SV40 DNA<sup>42</sup> in  $\lambda$ FIXII according to standard procedures. Cosmid and phage DNA was extracted using standard techniques involving purification over Qiagen tips (Qiagen). Positive cDNA clones were identified in a mixed poly-dT/random-primed library in  $\lambda$ gt11

constructed from human fetal kidney (Clontech). Library screening was performed by plaque hybridization using various DNA probes, derived from subsequently isolated cDNA clones, according to the manufacturer's instructions.

#### **DNA sequencing and computer analysis**

Nucleotide sequences were determined according to the dideoxy chain termination method using the T7 polymerase sequencing kit of Pharmacia/LKB or the dsDNA Cycle Sequencing System (GIBCO/BRL). Sequencing results were analyzed using an A.L.F. DNA sequencer<sup>TM</sup> (Pharmacia Biotech) on standard 30 cm, 6% Hydrolink<sup>R</sup>, Long Range<sup>TM</sup> gels (AT Biochem). Sequence analysis utilized Lasergene (DNASTAR) and BLAST and BEAUTY searches (NCBI).

#### **PCR amplification of genomic DNA**

PCR amplifications were carried out essentially as described before<sup>42</sup>. The following amplimers were used to generate a *PLAG1* exon 1 probe (5'-CAA TGG CTG CTG GAA AGA GG-3' and 5'-CCC GTC CGC CGC CTC TAC ACC-3'), a *PLAG1* ORF probe (5'-CGT AAG CGT GGT GAA ACC AAA C-3' and AGG GTC GTG TGT ATG GAG GTG A-3'), a *PLAG1* 3'-UTR probe (5'-ACA TGG CAT TTC GTG TCA CT-3' and 5'-CCA CAA TGG CTC TAG AT-3') and a *CTNNB1* exon 1 probe (5'-TGT GGC AGC AGC GTT GGC CCG GC-3' and 5'-CTC AGG GGA ACA GGC TCC TC-3').

#### **Rapid amplification of cDNA ends (RACE)**

Rapid amplification of 3' cDNA-ends (3'-RACE) was performed using a slight modification of part of the GIBCO/BRL 3'-ET protocol. For first strand cDNA synthesis, adapter primer (AP2) AAG GAT CCG TCG ACA TC(T)17 was used. For both initial and secondary rounds of PCR, the universal amplification primer (UAP2) CUA CUA CUA CUA AAG GAT CCG TCG ACA TC was used as "reversed primer". In the first PCR round the following specific "forward primers" were used: i) 5'-CAA TGG CTG CTG GAA AGA GG-3' (exon 1) or ii) 5'-AGA ATT TGG GCC TCA GAC AAG ATA-3' (3'-UTR, exon 5). In the second PCR round the following specific forward primers (nested primers as compared to those used in the first round) were used: i) 5'-CAU CAU CAU CAU GGC CGG AGG GAG GAT GTT AA-3' (exon 1) or ii) 5'-CAU CAU CAU CAU ATT GTC CTG GGT TGA TTA TGC AT-3' (3'-UTR, exon 5). CUA/CAU-tailing of the nested, specific primers allowed the use of the directional CloneAmp cloning system (GIBCO/BRL).

For 5'-RACE experiments, the Marathon cDNA Amplification kit (Clontech) was used according to the manufacturer's instructions with minor modifications. The 5'-untranslated end of the normal *PLAG1* transcript as well as the chimeric transcripts were isolated by 5'-RACE. First strand placenta or adenoma cDNA respectively, was

synthesized from 5 µg total RNA using the MV2 primer (5'-CTG CAC TGT ACC CAC CCC TGT GAT-3') located in exon 5. The ds cDNA was ligated to the adaptor and amplified using the anchor primer AP1 and the MV5 primer (5'-CAG GAG AAT GAG TAG CCA TGT GC-3') also located in exon 5. A second round of PCR was performed using the anchor primer and the MV6 primer (5'-TGC ACT TGT AGG GCC TCT CTC CTG-3') located in exon 4. The final PCR products were purified out of agarose gel and cloned into the pCRII vector (Invitrogen).

#### RT-PCR

Total RNA (5 µg) was reverse-transcribed using Superscript II reverse transcriptase (GIBCO BRL) and oligo d(T) primers according to the recommended conditions. 0.25 µg of the resulting cDNA was subject to amplification using a variety of primer sets. The amplification conditions for the *CTNNB1/PLAG1* fusion transcripts were 30 cycles at 94 °C for 10 sec and 68 °C for 1 min in a final volume of 50 µl using the Expand long template PCR system (Boehringer Mannheim). The first round PCR was carried out with the *CTNNB1* primer 5'-TGT GGC AGC AGC GTT GGC CCG-3' (CAT-UP) and the *PLAG1* primer 5'-CAG GAG AAT GAG TAG CCA TGT GC-3' (MV5). The second round was performed on a 20 fold diluted sample with the *CTNNB1* primer 5'-ACG GAG GAA GGT CTG AGG AGC AG-3' (NECAT-UP) and the *PLAG1* primer 5'-TGC ACT TGT AGG GCC TCT CTC CTG-3' (MV6). To amplify the reciprocal *PLAG1/CTNNB1* fusion transcript two rounds of PCR amplification were performed with 30 cycles at 94 °C for 30 sec, 63 °C for 30 sec and 72 °C for 1 min in a final volume of 50 µl. The first round was carried out with the *PLAG1* primer 5'-CAA TGG CTG CTG GAA AGA GG-3' (START-UP) and the *CTNNB1* primer 5'-AAG GAG CTG TGG TAG TGG CAC-3' (CAT3). The second round was performed on a 20 fold diluted sample with the *PLAG1* primer 5'-GGC CGG AGG GAG GAT GTT AA-3' (START-RACE) and the *CTNNB1* primer 5'-GCC GCT TTT CTG TCT GGT TCC A-3' (CAT3NEST).

## ACKNOWLEDGEMENTS

The authors thank M. Behrendt for skilful cytogenetic analysis and acknowledge the excellent technical assistance of C. Huysmans, J. Geurts, R. Thoelen and M. Willems. We thank Dr. F. Speleman for providing us with tumor T9587. This work was supported in part by the Flanders Interuniversity Institute for Biotechnology, by the European Union through Biomed 1 program "Molecular Cytogenetics of Solid Tumours", the "Geconcerteerde Onderzoekacties 1992-1996", the "Nationaal Fonds voor Wetenschappelijk Onderzoek" (NFWO), the "ASLK-programma voor Kankeronderzoek", the Swedish Cancer Society and the IngaBritt and Arne Lundbergs Research Foundation. K. Kas is a post-doc of the NFWO. M. Voz is a post-doc of "Korn op tegen Kanker" (NFWO).

## REFERENCES

1. Schoenmakers, E.F.P.M. *et al.* Recurrent rearrangements in the high mobility group protein gene, *HMGIC*, in benign mesenchymal tumours. *Nature Genetics* **10**, 436-444 (1995).
2. Ashar, H.R. *et al.* Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains. *Cell* **82**, 57-65 (1995).
3. Bustin, M., Lehn, D.A. & Landsman, D. Structural features of the HMG chromosomal proteins and their genes. *Biochim. Biophys. Acta* **1049**, 231-243 (1990).
4. Wolffe, A.P. Architectural transcription factors. *Science* **264**, 1100-1101 (1994).
5. Petit, M.P.M., Mols, R., Schoenmakers, E.F.P.M., Mandahl, N. & Van de Ven, W.J.M. *LPP*, the preferred fusion partner gene of *HMGIC* in lipomas is a novel member of the LIM protein gene family. *Genomics*, In press, (1996).
6. Schmeichel, K.L. & Beckert, M.C. The LIM domain is a modular protein-binding interface. *Cell* **79**, 211-219 (1994).
7. Seifert, G., Brocheriou, C., Cardesa, A., & Eveson, J.W. WHO international histological classification of tumours: tentative histological classification of salivary gland tumours. *Pathol. Res. Pract.* **186**, 555-581 (1990).
8. Mitelman, F. Catalog of Chromosome Aberrations in Cancer. 4th ed., New York, Wiley- Liss (1994).
9. Mark, J., Dahlenfors, R., Ekedahl, C. & Stenman, G. The mixed salivary gland tumor - a normally benign human neoplasm frequently showing specific chromosome abnormalities. *Cancer Genet. Cytogenet.* **2**, 231-241 (1980).
10. Peifer, M. Cancer, catenins, and cuticle pattern: a complex connection. *Science* **262**, 1667-1668 (1993).
11. Røijer, E. *et al.* Identification of a YAC spanning the chromosome 8q12 translocation breakpoint in pleomorphic adenomas with t(3.8)(p21;q12). *Genes Chromosom. Cancer*, In press.
12. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
13. Bellefroid, E.J. *et al.* The human genome contains hundreds of genes coding for finger proteins of the Kruppel type. *DNA* **8**, 377-387 (1989).
14. Sachs, A.B. Messenger RNA degradation in eukaryotes. *Cell* **74**, 413-421 (1993).

15. Nollet, F., Berx, G., Molemans, F. & van Roy, F. Genomic organization of the human  $\beta$ -catenin gene (CTNNB1). *Genomics* **32**, 413-424 (1996).
16. Kraus, C. *et al.* Localization of the human  $\beta$ -catenin gene (CTNNB1) to 3p21: a region implicated in tumor development. *Genomics* **23**, 272-274 (1994).
17. Gemmill, R.M. *et al.* A second-generation YAC contig map of human chromosome 3. *Nature* **377**, 299-319 (1995).
18. Rabbitts, T.H. Chromosomal translocations in human cancer. *Nature* **372**, 143-149 (1994).
19. Ye, B.H. *et al.* Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. *EMBO J.* **14**, 6209-6217 (1995).
20. Dallery, E. *et al.* TTF, a gene encoding a novel small G protein, fuses to the lymphoma-associated LAZ3 gene by the t(3;4) chromosomal translocation. *Oncogene* **10**, 2171-2178 (1995).
21. Galiegue-Zouitina, S. *et al.* The B cell transcriptional coactivator BOB1/OBF1 gene fuses to the LAZ3/BCL6 gene by t(3;11)(q27;q23.1) chromosomal translocation in a B cell leukemia line (Karpas 231). *Leukemia* **10**, 579-587 (1996).
22. Aplan, P.D. *et al.* Involvement of the putative hematopoietic transcription factor SCL in T-cell acute lymphoblastic leukemia. *Blood* **79**, 1327-1333 (1992).
23. El-Baradi, T. & Pieler, T. Zinc finger proteins: what we know and what we would like to know. *Mech. Develop.* **35**, 155-169 (1991).
24. Haber, D.A. *et al.* An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell* **61**, 1257-1269 (1990).
25. Ladanyi, M. & Gerald, W. Fusion of the EWS and WT1 genes in the desmoplastic round cell tumor. *Cancer Res.* **54**, 2837-2840 (1994).
26. De The, H. *et al.* The PML-RAR fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell* **66**, 675-684 (1991).
27. Chen, Z. *et al.* Fusion between a novel Kruppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. *EMBO J.* **12**, 1161-1167 (1993).
28. Kerckaert, J.-P. *et al.* LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. *Nature Genetics* **5**, 66-70 (1993).

29. Ye, B.H. *et al.* Alterations of a zinc finger-encoding gene, ECL-6, in human large-cell lymphoma. *Science* **262**, 747-750 (1993).
30. Mitchell, P.J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378 (1989).
31. Gumbiner, B.M. & McCrea, P.D. Catenins as mediators of the cytoplasmic functions of cadherins. *J. Cell Sci.* **17**, 155-158 (1993).
32. Kemler, R. From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends Genet.* **9**, 317-321 (1993).
33. Gumbiner, B.M. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**, 345-357 (1996).
34. Rubinfeld, B. *et al.* Association of the APC gene product with  $\beta$ -catenin. *Science* **262**, 1731-1734 (1993).
35. Su, L.-K., Vogelstein, B. & Kinzler, K.W. Association of the APC tumor suppressor protein with catenins. *Science* **262**, 1734-1737 (1993).
36. Polakis, P. Mutations in the APC gene and their implications for protein structure and function. *Curr. Opin. Genet. Dev.* **5**, 66-71 (1995).
37. Rubinfeld, B. *et al.* Binding of GSK3 $\beta$  to the APC- $\beta$ -catenin complex and regulation of the complex assembly. *Science* **272**, 1023-1026 (1996).
38. Peifer, M. Regulating cell proliferation: as easy as APC. *Science* **272**, 974-975 (1996).
39. Evans, R.W. & Cruickshank, A.H. Epithelial tumours of the salivary glands. *Major Probl. Pathol.* **1**, 1-299 (1970).
40. Dardick, I., Stratis, M., Parks, W.R., DeNardi, F.G. & Kahn, H.J. S-100 protein antibodies do not label salivary gland myoepithelium. Histogenetic implications for salivary gland tumors. *Am. J. Pathol.* **138**, 619-628 (1991).
41. Sandros, J., Stenman, G. & Mark J. Cytogenetic and molecular observations in human and experimental salivary gland tumors. *Cancer Genet. Cytogenet.* **44**, 153-167 (1990).
42. Schoenmakers, E.F.P.M. *et al.* Physical mapping of chromosome 12q breakpoints in lipoma, pleomorphic salivary gland adenoma, uterine leiomyoma, and myxoid liposarcoma. *Genomics* **20**, 210-222 (1994).
43. Sambrook, J., Fritsch, E.F. & Maniatis, T. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor (1989).
44. Feinberg, A.P. & Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6-13 (1984).

45. Albertsen, H.M. et al. Construction and characterization of a yeast artificial chromosome-library containing seven haploid human genome equivalents. *Proc. natl. Acad. Sci. U.S.A.* **87**, 4256-4260 (1990).
46. Green, E.D. & Olson, M.V. Systematic screening of yeast artificial-chromosome libraries using the polymerase chain reaction. *Proc. natl. Acad. Sci. U.S.A.* **87**, 1213-1217 (1990).
47. Wood, S., Schertzer, M., Drabkin, H., Patterson, D. & Deaven, L.L. Characterization of a human chromosome 8 cosmid library constructed from flow-sorted chromosomes. *Cytogenet. Cell Genet.* **59**, 243-247 (1992).

**Figure 1**

**A:** Contig of 3 overlapping YACs (bold lines), 27 cosmids and 2 phages, containing 27 landmarks and spanning a 300 kb DNA region on chromosome 8q12. Contig elements are labelled or numbered and defined in the list below. Cosmid clones isolated from the arrayed chromosome 8-specific cosmid library constructed at Los Alamos National Laboratory (LANL)<sup>47</sup> are named after their unique microtiter plate addresses. #1 and #22 are genomic phage clones; #21 is a clone isolated from the non-arrayed LANL chromosome 8-specific cosmid library. The orientation of the contig on the long arm of chromosome 8 is given as well as the order of 27 landmarks. It should be noted that the contig is not scaled. Below the contig, the genomic organization and the relative location of the *PLAG1* gene is given schematically, with exact sizes (bp) of its exons and estimated sizes (kb) of its introns. Noncoding sequences are represented as open boxes and coding sequences as black boxes. The relative positions of the translation initiation (ATG) and stop (TAG) codons in *PLAG1* are indicated. At the bottom of the figure, characteristics of the deduced protein encoded by *PLAG1* are given. Zinc fingers are labelled F1-F7.

1 = PEM18 + CH33 = D8SXXX	11 = 163E1 + CH280 = D8SXXX	21 = CEM55 + EM195 = D8SXXX
2 = 96E9 + PENK = D8SXXX	12 = 203B8 + CH280 = D8SXXX	22 = PEM49 + EM317 = D8SXXX
3 = 41E12 + PENK = D8SXXX	13 = 50B7 + EM187 = D8SXXX	23 = 149G12 + D8S285
4 = 17B6 + EM172 = D8SXXX	14 = 115H9 + EM187 = D8SXXX	24 = 64C1 + D8S285
5 = 93B2 + EM172 = D8SXXX	15 = 129B10 + CH129 = D8SXXX	25 = 64D1 + D8S285
6 = 47D11 + CH280 = D8SXXX	16 = 116A2 + CH129 = D8SXXX	26 = 143B1 + MOS = D8SXXX
7 = 53B10 + CH280 = D8SXXX	17 = 7B7 + CH129 = D8SXXX	27 = 143H1 + MOS = D8SXXX
8 = 66E5 + CH280 = D8SXXX	18 = 4H4 + EM156 = D8SXXX	28 = 148E9 + MOS = D8SXXX
9 = 144A7 + CH280 = D8SXXX	19 = 4H5 + EM156 = D8SXXX	29 = 22B8 + MOS = D8SXXX
10 = 163D11 + CH280 = D8SXXX	20 = 4H6 + EM156 = D8SXXX	

(clone number, cosmid or phage clone, STS, accession number)

**B:** Mapping of the 8q12 translocation breakpoint in an adenoma (CG644) with a t(3;8)(p21;q12). Cosmid CEM48 (green signal) was co-hybridized with an alpha-satellite probe specific for chromosome 8 (red signal). Hybridization signals were found on the normal 8, the der (3), and the der (8), indicating that CEM48 spans the t(3;8)(p21;q12) breakpoint. Chromosomes are stained in blue with DAPI.

**Figure 2**

**A:** cDNA and deduced amino acid sequence of *PLAG1*. The relative positions of exon/intron boundaries are indicated by triangles (▼). The conserved C, F, L and H residues in the zinc finger domains are underlined. Residues

22-25 and 29-32 constitute two potential nuclear localization signals. A potential polyadenylation signal and the putative mRNA destabilizing ATTTA motifs in the 3'-noncoding region are underlined. The nucleotide sequence of the complete cDNA has been deposited at GenBank under accession number XXX.

**B:** Alignment of the seven zinc-finger-like motifs found in the deduced amino acid sequence of *PLAG1* relative to the C2H2 consensus motif. The canonical C, F, L and H residues are given in bold.

### Figure 3

Illustrative example of the detection of DNA rearrangement in *PLAG1* of a primary pleomorphic adenoma using Southern blot analysis. DNA of pleomorphic adenoma CG368 (panel I) and control DNA isolated from normal lymphocytes (panel II) were digested with restriction endonuclease BamHI (B), EcoRI (E), HindIII (H), or PstI (P), as indicated. A probe with specificity for exon 3 of *PLAG1* (EM440) was used. The molecular weight markers are 23.1, 9.4, 6.5, 4.3, 2.3 and 2.0 kb, respectively.

### Figure 4

**A:** Detection of *CTNNB1/PLAG1* and *PLAG1/CTNNB1* fusion transcripts by RT-PCR in primary adenomas. I. *CTNNB1/PLAG1* were detected using the RT-PCR protocol and primers described in detail in Methods. Primary tumors analyzed included CG368 (lane 1), CG588 (lane 2), CG644 (lane 3), CG682 (lane 4), CG752 (lane 5), CG753 (lane 6), T9587 (lane 7), and CG580 (lane 8). II. *PLAG1/CTNNB1* fusion transcripts were detected similarly using the same samples as under "I". Details of the primers used here are given in Methods Section. PCR products are labelled A-D.

**B:** Schematic representation of the nature and origin of *CTNNB1/PLAG1* and *PLAG1/CTNNB1* fusion transcripts in primary pleomorphic adenomas with t(3;8)(p21;q12). At the top of the figure, the exon/intron distribution of the *PLAG1* gene is given, at the bottom, the exon/intron distribution for the *CTNNB1* gene. Positions of chromosome breakpoints are indicated by an arrow (↓). Translation initiation sites are indicated by asterisks (\*) and stop codons by triangles (▼). A-D: schematic exon compositions of hybrid transcripts, as established by 5'-RACE analysis. A: cDNA sequence junction between exon 1 of *CTNNB1* and exons 3-5 of *PLAG1*. B: cDNA sequence junction between exon 1 of *CTNNB1* and exons 2-5 of *PLAG1*. C: cDNA sequence junction between exon 1 of *PLAG1* and exons 2-16 of *CTNNB1*. D: cDNA sequence junction between exons 1-2 of *PLAG1* and exons 2-16 of *CTNNB1*.

#### Figure 5

A: Northern blot analysis of the expression pattern of *PLAG1* in normal human fetal tissues including brain (1), lung (2), liver (3), kidney (4) as well as adult tissues including heart (5), brain (6), placenta (7), lung (8), liver (9), skeletal muscle (10), kidney (11), and pancreas (12).

B: Northern blot analysis of the expression pattern of the *CTNNB1* gene in normal human fetal and adult tissues as described under "A".

C: Detection of *CTNNB1/PLAG1* transcripts in pleomorphic adenomas by Northern blot analysis using exon 1 of *CTNNB1* as a molecular probe. Lane 1, RNA of CG644 (t(3;8)), lane 2, RNA of CG580 (t(8;15)), and lane 3, RNA of CG682 (ins 3p21 (8q12)). The *CTNNB1/PLAG1* fusion transcript is indicated.

D: Using the same blot as under "C", detection of *CTNNB1/PLAG1* transcripts using a probe with specificity for the 3' UTR of *PLAG1* (probe KK64). The *CTNNB1/PLAG1* transcript is indicated.

#### Figure 6.

STSs used to generate the 300 kb cosmid contig at chromosome 8q12 encompassing the *PLAG1* gene. Some STSs marked with an asterisks to indicate that their relative locations in the contig shown in Fig. 1. are given.

#### Figure 7.

Nucleotide sequence of the *CTNNB1* cDNA.

#### Figure 8.

YAC contig of region of chromosome 8q12 encompassing the *PLAG1* gene. YAC clones were isolated from the CEPH mark 1 and mark 3 YAC libraries except for 900g10157 which was isolated from an ICRF YAC library. The relative positions of published STSs and genes as well as those corresponding to new STSs are indicated.

14 02 08 98

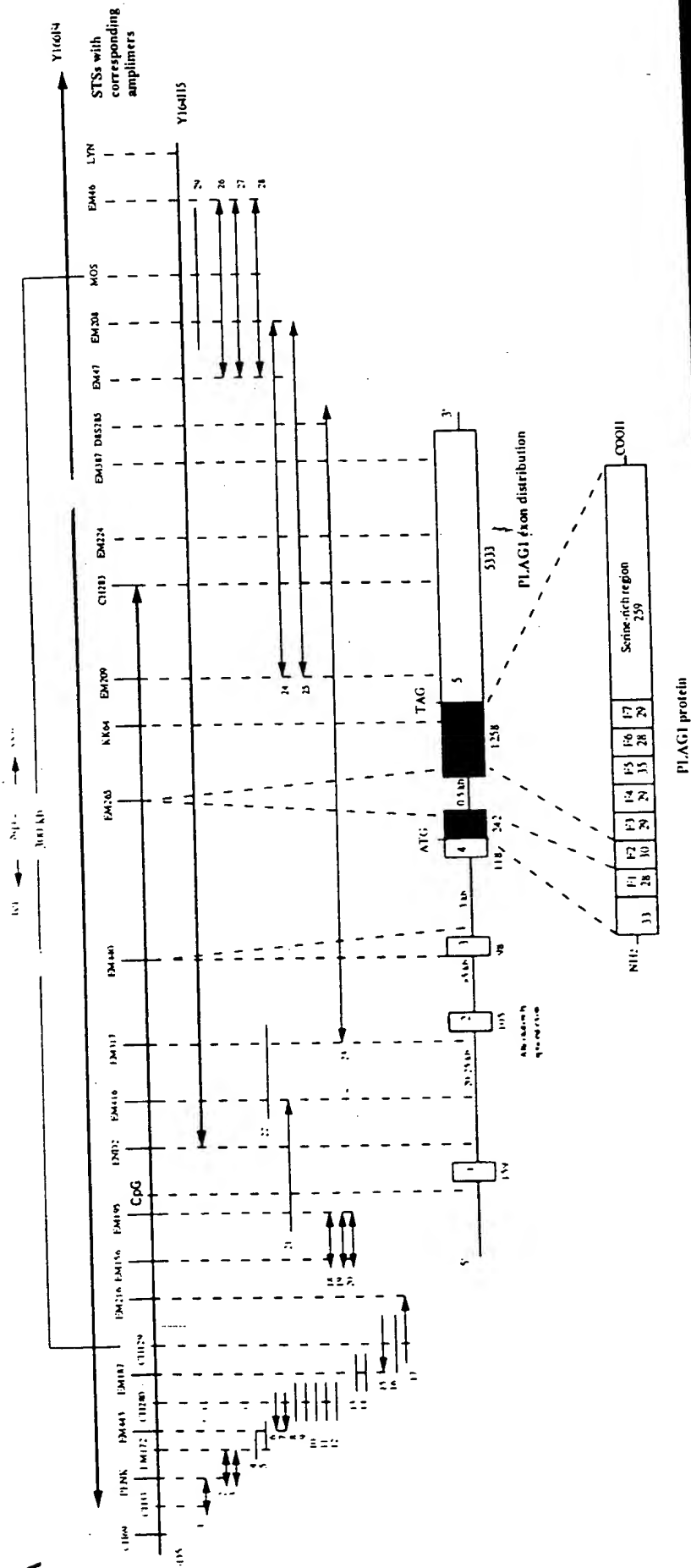
3

**CLAIM**

1. PLAG1-gene as described in the description.

## ABSTRACT

The largest cytogenetic subgroup of pleomorphic adenoma of the salivary glands carries chromosome 8q12 aberrations with 3p21 as preferential translocation partner. Here we demonstrate that the t(3;8)(p21;q12) results in promoter swapping between *PLAG1*, a novel, developmentally regulated zinc finger gene on 8q12, and the constitutively expressed gene for  $\beta$ -catenin (*CTNNB1*), a protein interface functioning in adherens junctions and the WG/WNT signalling pathway. Fusions occur in the 5'-non-coding parts of both genes, exchanging regulatory control elements while preserving the coding regions. Due to the t(3;8)(p21;q12), *PLAG1* is activated and expression of *CTNNB1* down-regulated. Activation of *PLAG1* was also observed in an adenoma with a variant translocation t(8;15). The results indicate that *PLAG1* activation due to promoter swapping is a crucial event in salivary gland tumorigenesis.



PLAG1 protein

B

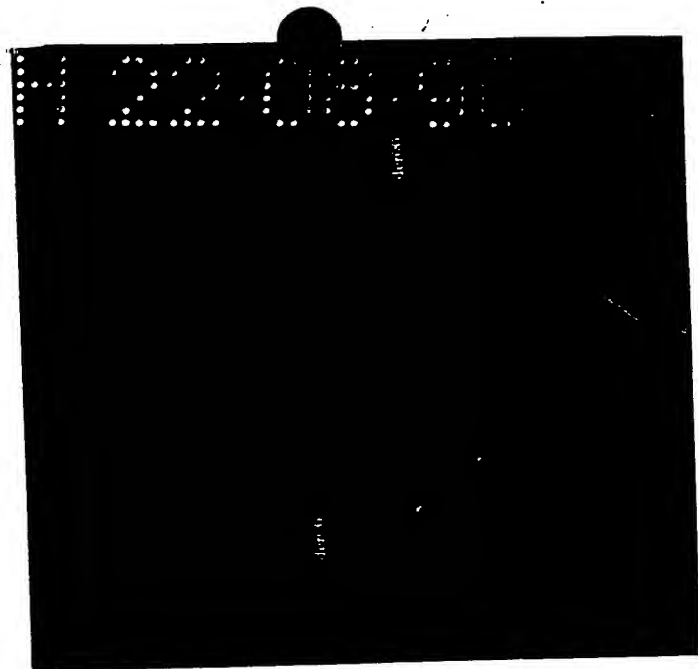


Figure 1



[illegible]

**Figure 2 A**

4/74

4.22.08.08

PLAG1 Finger 1 FPC..QLCDKAFNSVEKLVHSYS.HTGERP  
PLAG1 Finger 2 YKCIQQDCTKAFVSKYKLQRHMAT.HSPEKT  
PLAG1 Finger 3 HKC..NYCEKMFHRKDHLKNHLHT.HDPNKET  
PLAG1 Finger 4 FKCEE..CGKNYNTKLGFKRHLAL.HAATSGD  
PLAG1 Finger 5 LTC..KVCLQTFESTGVILLEHLKS.HAGKSSGGVKEKK  
PLAG1 Finger 6 HQCEH..CDRRFYTRKDVRRHMVV.HTGRKD  
PLAG1 Finger 7 FLC..QYCAQRFGRKDHLTRHMKKSHNQELL

PLAG1 Consensus ..C....C...F.....L..H....H.....

C2H2 Consensus FxCxxxxCxxxFxxxxxLxxHxxxxHxxxxx  
Y

5/14

100-100-00

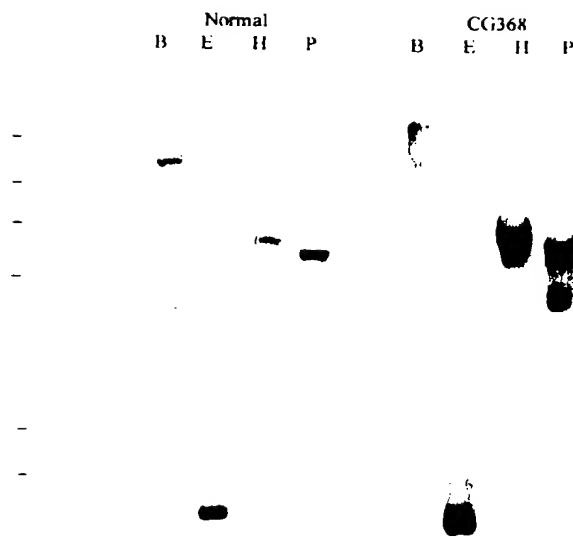


Figure 3

6174

100000

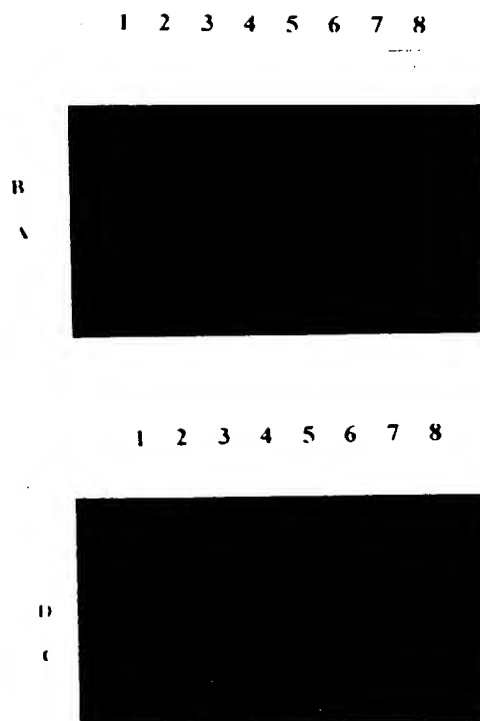


Figure 4a

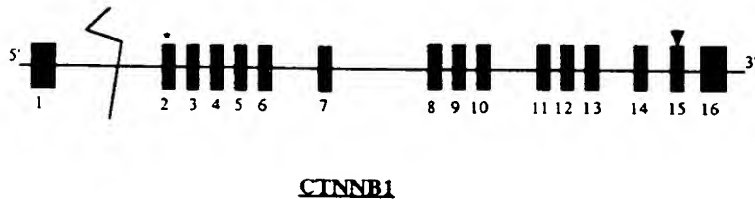
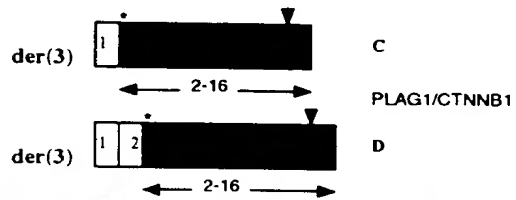
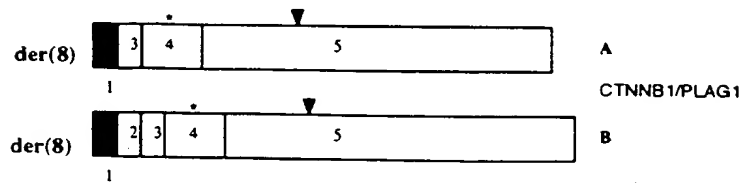
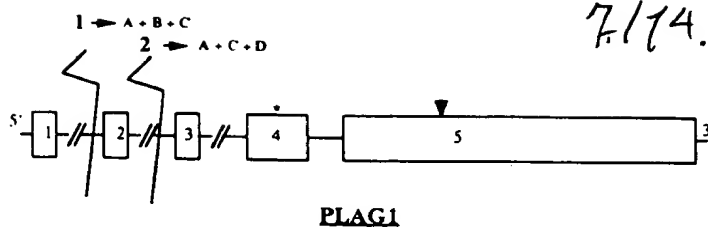


Figure 4 B

8/14

12.08.95

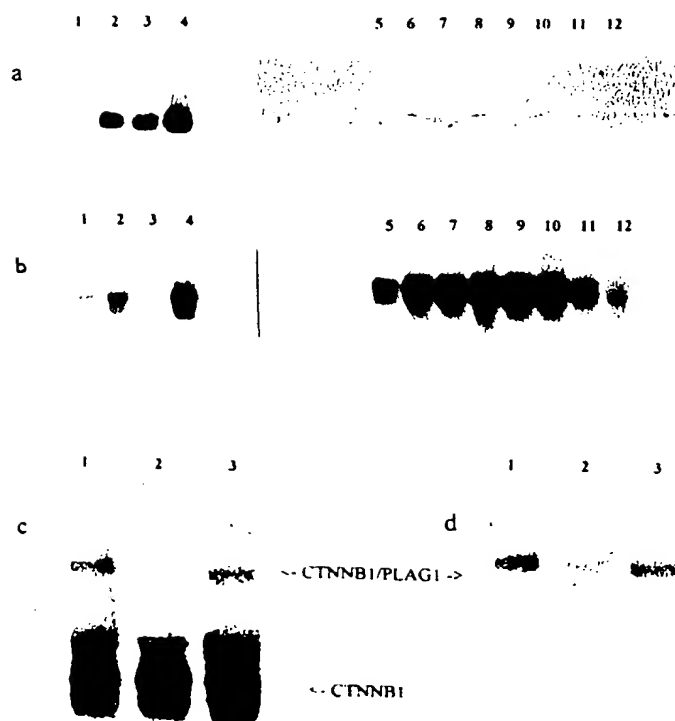


Figure 5

9/14

11010005

STSs used to generate the 300 kb cosmid contig mapping  
at chromosome 8q12 and encompassing PLAG1

STS CH129

GAATTCTAAAACCATTTTATAAATCATACTGAATCCCAGAACAATATATTTTAAACAACTTAA  
AAAAAAGAACAAAATAAAATAGCAAAACATTTTTAAGAGTGTAGATTCTTTGAAATTAAAGG  
ACATACTTACCCTGTAGT

STS CH280

GAATTCTTGCACCGGTTTTTTTCTTATCAGTGTGGGCTGATGTTCCATTAACTGTGGTGTAAAT  
TTGAGTATAGTCACTGACTGATTCTAGATATTTTCAGAGGGTCAAGACTTTTTCTAAGACCT  
TTATATGTGGTTGAATTCTTGTCTTGGTTTCACAGAAGGTATATTAGCAAAGCATTTTTGG  
TGTTGAAGCTTGGTCTGTGATCTAGT

STS CH33

GAATTCGTTTTTTATTTGACAAGCACATGAAGCCTTATCAGACGGAGGCCTCAATCCTTTGGC  
TGGGGTTTATAAGCAGGTAGCGCTAGACCTTCCCATTCTACATAAGCTGATGGGCACGGTAA  
TAGCTGGGGGTTTTCTCACAAGTCAAAGACAAATTGTCTGTTTTCAAGCGTGTGAAACAGTT  
WAAWACGTTTGAGGTCTCTCTCTTGCTTCATAGGCCATCTTGGCTCAGACATTCTACAGMCA

STS EM156

TCTGAGCAACAAGAGCGAAACTCCATCTCAAAATATATATATATATAGGTAATTGTTGTCAT  
TAATATTAATGTAGTAGCAGCAGCAACAGTCATGGTAGCAATATTGCTCTATTTGGGAGGCA  
ACTTATAATTATTAAGTGTGGAATATCTTTGAAAAATGTTTTTNGCAGAMGTTATGTTCCCA  
TTCCTGACTGGMGCTCATTATAAATACCCATCTTCTCTGAATAGCGCAAGGACTTTTGAAAA  
AGTGTCTGAGTAAAC

STS EM195

ACAATCAATTTTAGAAAGTAATCATTTTCATTACCCCCAAACTGAAACCCTGTACCTGTTAGCA  
CTCACTCCCCCTTTTCATTTTACTTTTTATTTATTTTATTTTGTGAGAGAGACTTGCTCTATC  
GCCCNGGCNVCAAGTGCAGTGGCACAACTCAACTCACTGCAACCTCTGCCTGCCAGGGTCAA  
GTGATTCTTGTGCCTCAGAGTCCCAAGTACCTGGGATTACAGGCATAAGCCACCACGCCTGG  
CTAAATTTTGTATTTTTCAGTAGTGACGGGGTTTCACCATGTTGGCCAGGCTGTCTCAAAGT  
CTGACCTCAGGTAATCCACCCTCCTCAGCCTCCCAGAGTTCTGGGATTACAGGCGTGACACC  
GTGCCTGGCTCATTTTATTTTATAGAGATGAGATCTCACTCTKWTTGCCAGGCTTCAGTGC  
ATTGGCGTCATGATGGCTCACTGCAGGCTTCAGCTCCTGGCCTCAAAGCATCCTTCCGCCTCA

STS EM208

CTAGGCGACAGAGCAAGACTCTGTCTCAARGAAAAAaAAAAAaVRAAAAAATTACCAAAAC  
TGACTACAGAAAaVHGVARGGTTGAATAGCCTTACATTTGGVAAATAATTTTTTATTTATAAT  
TAAAGATATTTTTATAAAaVVTACTTAGGCCCATAGGCTTCACAGGTTAATTGTATTAAA  
TATTTAAGGAAAAAATAATACCAATCTTATTCATAGTCTTTCAGAAAATAGAGGCGTATCCA  
TTTTTCTAACTCATTTTAAGAAATAGCATCATTCTAATATCAAAAGCAAACAAGGMCATTGC  
AAAGAAGAAGGGGAAGAAGGAAGAGGAAGAGGAGGAGAAAGGGAAGCAGGAGATGGAGAAGA  
AGGAAGCCAGGTACAGTGCAATATTTCTCATGAACATAAACACAATTTTTTAAAAAGTATTAM  
CAGGCTGGGCTTGGTGGCTCTCACCCGTAATCCCAGCMCTTTGGAAGGCCCAAGGCGGGTGG  
GHCACAAGGTCAGGGGTTCGAG

Figure 6

10/14

4 2 0 8 9 5

**STS EM216**

TACTRACTGCTGTGCAGTTBTCcTGCAGTCAGTTCAGAGGTCATTTCTAACGTTGCACTAT  
GGGKCTATTTAATAGGTTTCCTAAAGAACAACATATCTCTTTAbAGTTACTCAGAGGGTAC  
ACAATGATGATGTCACACAATTAATTACCTATTAAGACTGAAATCCAGCAATGCATAGKGTG  
TGGACTTTACGCACATCCAGAAAAAGTTCTAGCACAAATTGTTTTHGTMtYATATATTTTCAG  
AAGCCATAGAAACACTATTAAAGCCCTCCCTAATCACTTAGGGATGCAAAaTCAATAT

**STS EM317**

GACCAACAAAGGCACACAAAGATTGGTTGCTTTCTGAAGAATCTAAAAATGGCATTGGGTAT  
AGGAGTTGGGGAAGCAAGTTGTATAGGCACCTACACTTAAGATAATTTGTCAATTATACAAA  
TAATTTTTAAAGTTTAAGCCCCTTTCTGACATGACACGTCCTATGGGTCCTTCACCCTTYtK  
KTCTCCTSCAGAGCTCCAGTCTGCCYYTTYTTKSTCTCTGAGCTCCAAAAMCAGTGAWTCCCC  
TGAAGTTACCTAGMCCCMCCATACAGTTTGTGACTCCCTAWMcGGGGGTACCyTCCCATGY  
CTGGCTAATAYTGABTYTTGTDACCCTGGCTTCTGTGTTACTACATTTGTTTTARTGGAAT  
TWATwAArGGGAAGCCTATCAA

**STS EM416**

GAGCAACTGAaCDNAGATTGGGTGAGGTAAGATGTGGGCTGCACAGGTGAGGCTGGAGAGGT  
GGGGAGTGCGTCCCAGTCGGGGGAGAAGAAGAAAAGGGCAGACTAGGGTAGAAATGCTTATW  
ACTcCTGTGACTGGAGCTGATGGTGTCTTAAGGAAAGTGGTGGGAAGGGAGGVCTGCAGAAA  
GGCAAGGCTGGAGTCGACTGAAGGCTGGAGAGCCACTGCTTTAACAAGTGTAMCTGGAGATG  
GAAGGGGCTGCAGGACAGGTCACTCAGCCAGTKGTGTGGARGCAATCTCACC

**STS EM443**

TTGATATTTGTTCTAACTCCACATTAACCTATTGACAAATACTCTAAATTGTAGCTACCATCT  
GTTACGTAGCTAGCAGGTACCCTAACAGCAATGGGTCAGCTTTTGAGTAGCGTTTCAACCAT  
GTTACCTCGAGTACGGTGTGGTGAGGCCAGACGCAGATGGAGAGAAAGAAACAGAATCGAGC  
ATTTCCATTTTGTTTTGCTCACAGTCCCCAGGGGCAACACAGCACAGCCTACAGGACCATG  
AAGGGGAGCACTGGGGTCACTCATGAAGCAGGGAGGTGCGGCCAGTGGTGGGGGgCCTTTAT  
GTGTTTTCTCAGGAAGGAATGGGCAAGGCAGGGTAAGCATGTTTCAGGACTGGTTAATTTGA  
ATAACTTCAGGGGGgCTCTAGGGCCTGgRGGCTGCCCTGGTTGTCTGGTACCyGGSCTG

**STS EM46**

ATATCAATCTTGGGTCTATGTATGTTTTTGCTTTTCCCcAGTGTTCCAGGCATGATGCTAAG  
GATATAGTGGATGATGAAATATATGCTTGCTGAATATGGGAATAAGAATTATTTTATGATCA  
GAHTTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTCAcNMaggCTVGTGTGCAGTGGCATGAT  
CTCAGCTCACWGCAACCTCTGVCTCCTGGGTTCAAGTGATT

**STS EM47**

GTAGAGACACACTAGGCATGCACAGACCAGTGCAGAATGAACAATATTTGTTACATGTGTAG  
TTCTTTATGGTTTACAAAACCTCTCCAGCCATTATCTTCTTTCAGCCTTATAAAAGACAGAG  
CATATTTTATTATCCTCATTTACCTWHTCTAGTAAGGCATTTTTTCTTTTTTCTTACTAGA  
GATATAAGGCTTAGGAAAAAAGTGAATACTACGATAAATGAATACTAGGAAAAGACATCACA  
ATCACAATTTATTAATATCAGAAAACAGDTTTTAAGAATAAAATWTTCAAWAARgAAA

Figure 6 (continued)

11/14

11-02-98

STS END2

TAATTTATCACTACGGAATTCTGTGCAGTGAGATCAAAGAGCTGTGTATGCCCATAATGTGA  
TTTACAGCCATTTTGTAAAACTGTAAAATACCTTAATATTCAATTTGGCTTAAGGTACAT  
TGAGGACTTCTGGTTGAAAATTACAGAGTGGTGAAGATTC

Known STSs

PENK

D8S285

MOS

STSs part of PLAG1

EM265

KK64

KK63/EM209

KK55/CH283

EM224

EM387

Figure 6 (continued)

12/14

12:08:08

Nucleotide sequence of cDNA of CTNNB1 ( $\beta$ -catenin)

```

1  aagcctctcg gtctgtggca gcagcgttgg cccggccccc ggagcggaga gcgaggggag
61  gcggagacgg aggaaggtct gaggagcagc ttcagtcccc gccgagccgc caccgcaggt
121 cgaggacggg cggactcccc cggcgggagg agcctgttcc cctgagggta tttgaagtat
181 accatacaac tgttttgaaa atccagcgtg gacaatggct actcaagctg atttgatgga
241 gttggacatg gccatgggaa cagacagaaa agcggctgtt agtcactggc agcaacagtc
301 ttacctggac tctggaatcc attctgggtg cactaccaca gctccttctc tgagtggtaa
361 aggcaatcct gaggaagagg atgtgggatac ctcccaagtc ctgtatgagt gggaacaggg
421 attttctcag tccttctactc aagaacaagt agctgatatt gatggacagt atgcaatgac
481 tcgagctcag agggtagcag ctgctatgtt ccctgagaca ttagatgagg gcatgcagat
541 cccatctaca cagtttgatg ctgctcatcc cactaatgtc cagcgttttg ctgaaccatc

601 acagatgctg aaacatgcag ttgtaaaactt gattaactat caagatgatg cagaacttgc
661 cacacgtgca atccctgaac tgacaaaact gctaaatgac gaggaccagg tgggtggttaa
721 taaggctgca gttatgggtc atcagcttcc taaaaaggaa gcttccagac acgctatcat
781 gcgttctcct cagatgggtg ctgctattgt acgtaccatg cagaatacaa atgatgtaga
841 aacagctcgt tgtaccgctg ggaccttgca taacctttcc catcatcgtg agggcttact
901 ggccatcttt aagtctggag gcattcctgc cctggtgaaa atgcttgggt caccagtgga
961 ttctgtgttg ttttatggca ttacaactct ccacaacctt ttattacatc aagaaggagc
1021 taaaatggca gtgcgttttag ctggtgggct gcagaaaatg gttgccttgc tcaacaaaac
1081 aaatgttaaa ttcttggtta ttacgacaga ctgccttcaa attttagctt atggcaacca
1141 agaaagcaag ctcatcatac tggctagtgg tggaccccaa gctttagtaa atataatgag

1201 gacctatact tacgaaaaac tactgtggac cacaagcaga gtgctgaagg tgctatctgt
1261 ctgctctagt aataagccgg ctattgtaga agctggtgga atgcaagctt taggacttca
1321 cctgacagat ccaagtcaac gtcttggttca gaactgtctt tggactctca ggaatcttcc
1381 agatgctgca actaaacagg aagggatgga aggtctcctt gggactcttg ttcagcttct
1441 gggttcagat gatataaatg tggtcacctg tgcagctgga attctttcta acctcacttg
1501 caataattat aagaacaaga tgatgggtctg ccaagtgggt ggtatagagg ctcttgtgctg
1561 tactgtcctt cgggctgggt acaggaaga cactactgag cctgccatct gtgctcttcg
1621 tcactctgacc agccgacacc aagaagcaga gatggcccag aatgcagttc gccttcaacta
1681 tggactacca gttgtgggta agctcttaca cccaacatcc cactggcctc tgataaaggc
1741 tactgttgga ttgattcgaa atcttgccct ttgtcccga aatcatgcac ctttgcgtga

1801 gcaggggtgcc attccacgac tagttcagtt gcttggttcgt gcacatcagg ataccacgag
1861 ccgtacgtcc atgggtggga cacagcagca atttgtggag ggggtccgca tggagaagaat
1921 agttgaaggt tgtaccggag cccttcacat cctagctcgg gatgttcaca accgaattgt
1981 tatcagagga ctaaatacca ttccattgtt tgtgcagctg ctttattctc ccattgaaaa
2041 catccaaaga gtagctgcag gggtcctctg tgaacttgct caggacaagg aagctgcaga
2101 agctattgaa gctgagggag ccacagctcc tctgacagag ttacttcaact ctaggaatga
2161 aggtgtggcg acatatgcag ctgctgtttt gttccgaatg tctgaggaca agccacaaga
2221 ttacaagaaa cggctttcag ttgagctgac cagctctctc ttcagaacag agccaatggc
2281 ttggaatgag actgctgac ttggacttga tattggtgcc cagggagaac cccttgagata
2341 tcgccaggat gatcctagct atcgttcttt tcactctggt ggatatggcc aggatgcctt

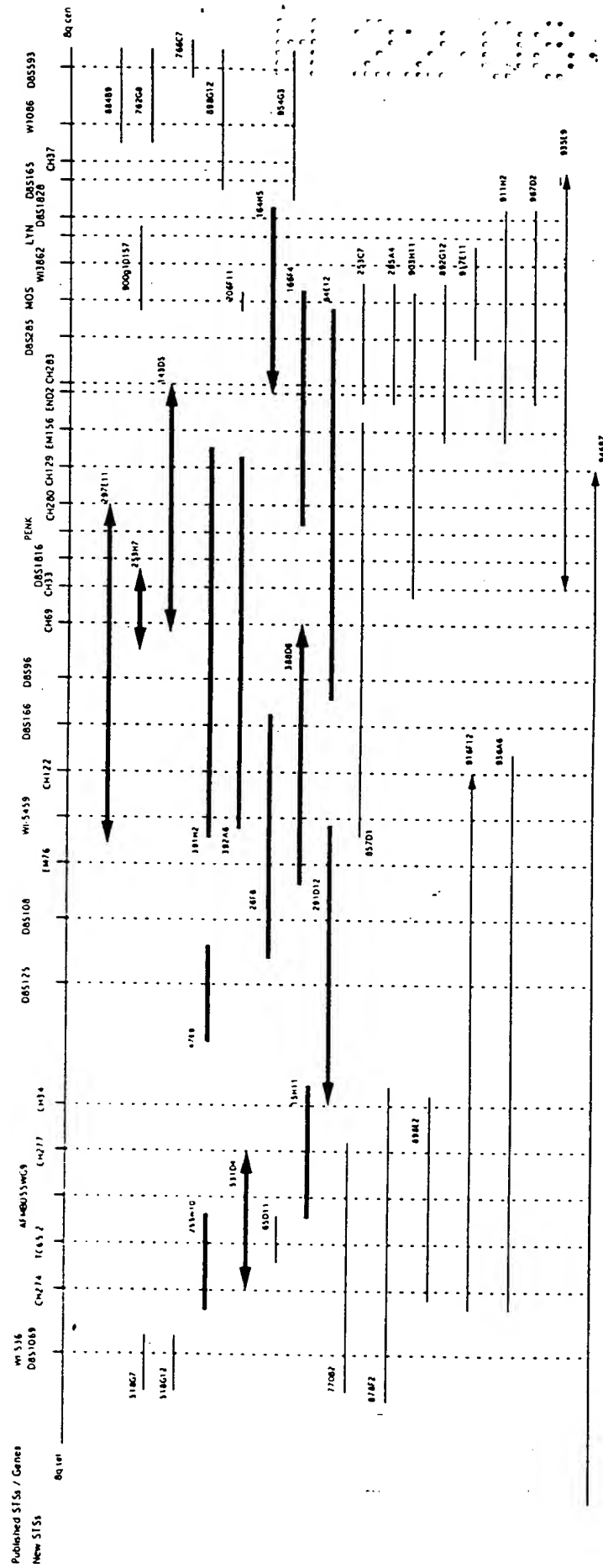
```

Figure 7

2401	gggtatggac	cccatgatgg	aacatgagat	gggtggccac	caccctgggtg	ctgactatcc
2461	agttgatggg	ctgccagatc	tggggcatgc	ccaggacctc	atggatgggc	tgcctccagg
2521	tgacagcaat	cagctggcct	ggtttgatac	tgacctgtaa	atcatccttt	agctgtattg
2581	tctgaacttg	cattgtgatt	ggcctgtaga	gttgctgaga	gggctcgagg	ggtgggctgg
2641	tatctcagaa	agtgcctgac	acactaacca	agctgagttt	cctatgggaa	caattgaagt
2701	aaactttttg	ttctggtcct	ttttggtcga	ggagtaacaa	tacaaatgga	ttttgggagt
2761	gactcaagaa	gtgaagaatg	cacaagaatg	gatcacaaga	tggaatttag	caaaccctag
2821	ccttgcttgt	taaaattttt	tttttttttt	ttttaagaat	atctgtaatg	gtactgactt
2881	tgcttgcttt	gaagtagctc	tttttttttt	tttttttttt	tttttttgca	gtaactgttt
2941	tttaagtctc	tcgtagtggt	aagttatagt	gaatactgct	acagcaattt	ctaattttta
3001	agaattgagt	aatgggtgtag	aacactaatt	aattcataat	cactctaatt	aattgtaatc
3061	tgaataaagt	gtaacaattg	tgtagccttt	ttgtataaaa	tagacaaata	gaaaatggtc
3121	caattagttt	ccttttttaat	atgcttaaaa	taagcagggtg	gatctatttc	atgtttttga
3181	tcaaaaaacta	tttgggatat	gtatgggtag	ggtaaatcag	taagagggtgt	tatttggaac
3241	cttgtttttg	acagttttacc	agttgccttt	tatcccaaag	ttgttgtaac	ctgctgtgat
3301	acgatgcttc	aagagaaaat	gcggttataa	aaaatgggtc	agaattaaac	ttttaattca
3361	tt					

Figure 7 (continued)

Figure 1 consists of three scatter plots arranged horizontally. Each plot has 'Number of children' on the x-axis and 'Number of mothers' on the y-axis. The first plot shows a positive correlation with a regression line. The second plot shows a positive correlation with a regression line. The third plot shows a positive correlation with a regression line.



**Figure 8**